Grand Challenge on Neural Decoding for Motor Control of non-Human Primates

Zhou, Biyan; Sun, Pao Sheng Vincent; Yik, Jason; Frenkel, Charlotte; Reddi, Vijay Janapa; Basu, Arindam

DOI
10.1109/BioCAS61083.2024.10798373

Publication date
2024

Document Version
Final published version

Published in
2024 IEEE Biomedical Circuits and Systems Conference, BioCAS 2024

Citation (APA)
Zhou, B., Sun, P. S. V., Yik, J., Frenkel, C., Reddi, V. J., & Basu, A. (2024). Grand Challenge on Neural Decoding for Motor Control of non-Human Primates. In *2024 IEEE Biomedical Circuits and Systems Conference, BioCAS 2024* (2024 IEEE Biomedical Circuits and Systems Conference, BioCAS 2024). IEEE. https://doi.org/10.1109/BioCAS61083.2024.10798373

Grand Challenge on Neural Decoding for Motor Control of non-Human Primates

Biyan Zhou
Electrical Engineering
City University of Hong Kong
Kowloon, Hong Kong
byzhou4-c@my.cityu.edu.hk

Pao-Sheng Vincent Sun
Electrical Engineering
City University of Hong Kong
Kowloon, Hong Kong
vincesun-c@my.cityu.edu.hk

Jason Yik
School of Engineering and Applied Sciences
Harvard University
Cambridge, USA
jyik@g.harvard.edu

Charlotte Frenkel
Department of Microelectronics
Delft University of Technology
Delft, Netherlands
c.frenkel@tudelft.nl

Vijay Janapa Reddi
School of Engineering and Applied Sciences
Harvard University
Cambridge, USA
vj@eecs.harvard.edu

Arindam Basu
Electrical Engineering
City University of Hong Kong
Kowloon, Hong Kong
arinbasu@cityu.edu.hk

2024 IEEE Biomedical Circuits and Systems Conference (BioCAS) | 979-8-3503-5495-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/BioCAS61083.2024.10798373

***Abstract*—To give paralyzed people hope for a normal life, Brain Machine Interfaces (BMI) record signals from the motor cortex and a decoder translates these "thoughts" to action. A high accuracy decoder is needed for a seamless user experience. At the same time it needs to be compact and low-power to support its integration in an implant to enable the compression required in wireless implantable BMIs. Hence, a model with a good trade-off between accuracy and resource requirement is desirable. In the IEEE BioCAS 2024 conference, we organized the first grand challenge on neural decoding for motor control. The evaluations were performed using the recently developed Neurobench software suite for benchmarking neuromorphic systems. There were two tracks –one preferring solutions with highest accuracy while the other gave weightage to the tradeoff between accuracy and implementation complexity. Out of the 10 teams registered for this event, the top 3 teams are invited to present their works in the IEEE BioCAS 2024.**

***Index Terms*—Grand challenge, machine learning, implantable BMIs, benchmarking, neuromorphic systems**

I. Organization of the Grand Challenge

A study reported in 2013 pegs the number of people who have paralysis at approximately 6 million in the US [1]. The need to restore their ability to perform activities for daily living has motivated the development of a host of assistive technologies. The most promising among the reported ones is the intra-cortical Brain-Machine Interface (iBMI). iBMIs aim to substantially improve patients' lives affected by spinal cord injury or debilitating neurodegenerative disorders such as tetraplegia, amyotrophic lateral sclerosis etc. These systems take neural activity as input and drive effectors such as a computer cursor [2], wheelchair [3], and prosthetic limbs [4] for communication, locomotion, and artificial hand control, respectively. Recently, there are also efforts to use similar technology to address speech disorders by converting imagined speech to produce a linguistic output [5].

We acknowledge a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11200922).

However, despite compelling advances, barriers to scalability and clinical translation remain. Apart from scaling the number of sensors [6], two major challenges related to system usability [7] are highlighted next. First, the decoding algorithms are currently being implemented on a PC with a wired connection from the headstage, increasing the bulkiness of the overall system and reducing patient mobility [2]– [4], [8]. We refer to this as the "mobility issue." Second, the present systems (we refer to them as wired iBMI or simply iBMI) involve the use of wires for data transmission from the implant through a hole in the skull, increasing the risk of infection. We refer to this as the "skull opening issue."

Wireless iBMIs can solve the two issues related to infection due to a skull hole and mobility due to wiring to a computer. However, this also raises scalability issues, as wireless iBMIs are difficult to scale beyond data rates of few tens of Mbps [9], [10] due to increased bit-error rates, low run-time between battery charges, as well as power dissipation constraints within cortical implants of 80 mW/cm2 [11], [12] leading to limiting the number of recording channels to around 100 (150-200 neurons). It is expected that dexterous prostheses would require simultaneous recording from 10,000 neurons [13]; a more refined understanding of the brain also requires recording an increasing number of neurons and is hence an essential goal in this field. Therefore, there is an urgent need to explore solutions to compress neural data to fit the wireless budget available for implants.

Several solutions have been proposed to compress the neural data. Compression schemes such as compressive sensing (CS) [14] or Autoencoder (AE) [15] fall short of the requirements to meet the available wireless data rates as the number of channels increases to 10,000. Only the integration of decoders (Dec) [[16], [17], [18]] in the implant can solve the problem in a scalable way. This underlines the importance of developing new neural decoders with a good tradeoff between accuracy and resource usage, suitable for deployment in implants.

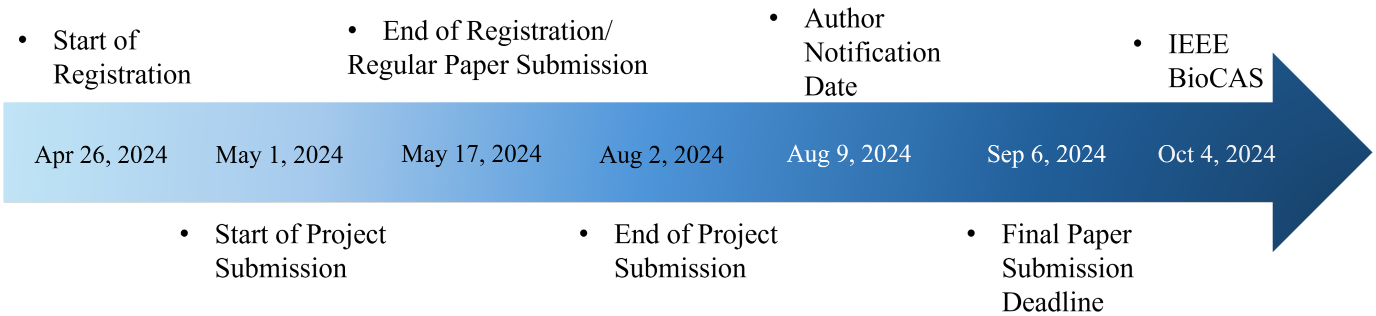

Fig. 1: Timeline of Grand Challenge on Neural Decoding

While traditional decoder designs have used signal processing approaches, recent advances in neural networks (such as model compression, quantization and new networks like spiking neural networks) are promising candidates for future iBMIs. This IEEE BioCAS Grand Challenge is geared in that direction and aims to push the boundary for neural decoder design toward next-generation BMI systems. Two tracks were included in the competition – one preferring solutions with highest accuracy while the other placed emphasis on the tradeoff between accuracy and implementation complexity.

10 teams worldwide registered for this competition to explore neural decoding algorithms and energy-efficient strategies. The registration phase started from from 26 April and continued to 17 May 2024. Aug 2 was the deadline for submission of solutions. The timeline of the grand challenge of neural decoding for motor control of non-human primates is summarized in Figure 1.

To ensure the high quality of the accepted papers, we manually checked the code submitted with the results written in their paper. The final ranking is determined based on the output results from their code and the quality of their papers. Finally, the top 3 teams are invited to present their work in the IEEE BioCAS 2024.

II. Summary of Participants

The IEEE BioCAS Grand Challenge has drawn in a total of 10 teams from 13 universities/organizations/institutions. The regional distribution of participants is presented in Table I. Finally, 5 solutions were found to satisfy all requirements of the competition. After peer review, 3 teams were selected as winners, one each in Track 1 and Track 2, and a 3rd team whose solution came after these two.

TABLE I: Region Distribution of Participants

Tracks	Region	Number of Teams
1	America	3
	Asia	3
	Europe	3
2	America	5
	Asia	4
	Europe	4

TABLE II: Number of reaches in each file

Filename	Number of Reaches
indy_20160622_01	970
indy_20160630_01	1023
indy_20170131_02	635
loco_20170210_02	587
loco_20170215_02	409
loco_20170301_05	472

III. Dataset

The dataset chosen for this challenge consists of recordings from the "Nonhuman Primate Reaching with Multichannel Sensorimotor Cortex Electrophysiology" dataset [19]. This dataset contains the recording of spikes generated by two Macaque monkeys while they were tasked to make self-paced reaches to targets placed in an 8×8 grid, without gaps or pre-movement delay intervals. One monkey reached with the right arm (recordings made in the left hemisphere of their brain) while the other reached with their left arm (recordings made in the right hemisphere of their brain). For most of the sessions, only the M1 (primary motor cortex) recordings were made over 96 channels, for the rest both M1 and S1 (somatosensory cortex) were recorded over 192 channels. The data from monkey 1 (named Indy) contains recording with 96 channels, while those from monkey 2 (named Loco) contains recording with 192 channels. The recordings contain 37 sessions recorded over 10 months for monkey 1 and 10 sessions over 1 month for monkey 2. We have carefully chosen six specific recordings from this dataset comprising three recordings from two monkeys each, with the recording dates spanning beginning, middle and end of the time of their respective total sessions. The details of the chosen files are listed in Table II.

The recordings are stored in Matlab's .mat file format, which contains seven variables in each file. Variables used for this challenge are cursor_pos, t and spikes.

- cursor_pos (k*2): The position of the cursor in Cartesian coordinates (x, y), expressed in millimeters. k refers to the number of samples in the file.
- t (k*1): The timestamp corresponding to each sample of the cursor_pos, expressed in seconds.

```python
import torch, snntorch
from torch.utils.data import DataLoader, Subset
from neurobench.datasets import PrimateReaching # dataset
from neurobench.models import TorchModel # model wrapper
from neurobench.benchmarks import Benchmark # runtime

dataset = PrimateReaching(path, file_name, ...)
loader = DataLoader(Subset(dataset, dataset.ind_test), ...)

net = torch.load(...) # trained model
model = TorchModel(net)
model.add_activation_module(snn.SpikingNeuron) # configure snn neuron

preprocessors = [] # none for this task
postprocessors = [] # none for this task
static_metrics = ["footprint"] # model-only measurements (size, connectivity, etc.)
workload_metrics = ["r2", "synaptic_operations"] # data-dependent measurements

benchmark = Benchmark(model, loader, preprocessors, postprocessors, [static_metrics, workload_metrics])
results = benchmark.run()
```

Listing 1: Example user-level interface for benchmarking an SNN with the NeuroBench code harness.

- spikes (n*u): A cell array of spike event vectors. Each element in the cell array is a vector of spike event timestamps, in seconds. The first unit (u1) is the "unsorted" unit, meaning it contains the threshold crossings which remained after the spikes on that channel were sorted into other units (u2, u3, etc.) For some sessions spikes were sorted into up to 2 units (i.e. u=3); for others, 4 units (u=5). n refers to number of recording channels and u refers to number of sorted units.

Every file is divided into training, validation, and test sets with a split ratio of 50% for training, 25% for validation, and 25% for testing.

IV. Code Harness

For this challenge, the data download and evaluation metrics are all encapsulated with the NeuroBench algorithm benchmarks [20], a community-driven project. Each submission used this common framework for loading data and benchmarking algorithms.

The harness features the task for this grand challenge, non-human primate motor prediction. In addition, it includes a suite of benchmarks of interest to neuromorphic algorithms, including event camera vision, continual learning, and sequence forecasting.

Given a trained model and an evaluation dataset, the harness automatically tests the model on the dataset, and calculates metrics at runtime. The metrics measure not only the correctness of the model, but also its compute costs. Listing 1 shows an example user-level interface for the code harness.

V. Grand Challenge

The grand challenge focuses on predicting finger movement velocities with a two-track competition.

A. Regression Tasks

Track 1: Obtaining highest accuracy as measured by the R2 metric Track 1 is a regression challenge aiming at finger motion velocity prediction using the spikes recorded from the multi-electrode array. The group with the highest accuracy is considered the winner for track 1.

Track 2: Obtaining the best tradeoff between accuracy and solution complexity The task of track 2 is similar to track 1, but accuracy is not the single consideration. The winner for track 2 should be the group with the best trade-off between prediction accuracy and model complexity.

B. Evaluation Criteria

The finger velocity prediction accuracy is measured using the R-Squared score. The equation is defined as Equation (1).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{1}$$

In this contest, R_X^2 and R_Y^2 are calculated separately, and the final result R^2 is the mean of the two. In addition, y_i and $\hat{y_i}$ are defined as the label and prediction respectively while $\bar{y}$ is the mean of labels.

Complexity is defined as the memory footprint and the synaptic operations (Dense/MACs/ACs). The memory footprint shows the model size in bytes. For synaptic operations, dense includes operations that concerns zeros. MACs denotes the total number of Multiply-Accumlates operations used for ANNs with continuous-valued activations. While ACs indicate the Accumlates operations, used for SNNs with binary activations.

VI. Summary of the Top 3 Winning Teams

The solutions of the top-3 teams are described here. Apart from studying the impact of various kinds of artificial neural networks on iBMI systems, this section will also involve examining various structures of spiking neural networks, data augmentation, data preprocessing methods, weight pruning techniques, efficient loss functions, and gradient propagation strategies to achieve an energy-efficient implementation. Table III reported the statistical results of top-3 team and Figure 2

TABLE III: Performance comparison of top-3 teams

Track	Rank	Team	Average R^2 score	Average Footprint (Bytes)	Average Dense	Average MACs	Average ACs
Track 1	1	BioCircuitBreakers	0.6982	45520	54283.2536	25316.4339	0
	2	QAAS-ZenkeLab	0.6978	4833360	1206272	0	42003.5272
	3	Primate Whisperers	0.6209	174104	4947.2461	627.2461	247.9486
Track 2	1	QAAS-ZenkeLab	0.6604	27144	13440	0	304.1519
	2	BioCircuitBreakers	0.6982	45520	54283.2536	25316.4339	0
	3	Primate Whisperers	0.6209	174104	4947.2461	627.2461	247.9486

illustrated the performance comparison between top-3 teams and [20].

A. Team BioCircuitBreakers

[21] from ETH Zurich and University of Zurich (Yuanxi Wang et al.) won first place in track 1 of this contest. They proposed an inference network structure called AEGRU: Two fully connected (FC) layers are positioned before and after the core, a recurrent neural network with a GRU layer. In addition, an Auxiliary training branch is used in training, which includes three FC layers and a firing rate reconstruction. The diverse network structures during training and inference guarantee a good trade-off between accuracy and complexity in inference, also enabling the extraction of latent features during training. Furthermore, they used the window summation method and the softplus function with logarithm for data preprocessing. The mean square error loss and Poisson negative log-likelihood loss are combined to make more accurate adjustments to model weights. Eventually, they have achieved a mean R^2 score of 0.6982 with a good trade-off in terms of complexity and accuracy.

B. Team QAAS-ZenkeLab

[22] from Zhejiang University (Tengjun Liu et al.) won first place in track 2 of this contest. They introduced two recurrent SNNs (RSNNs) for regression in motor decoding. One of the models is the bigRSNN with 1024 in a hidden layer and five readouts heads, which aims to achieve the highest R^2 score, while the other one is the tinyRSNN with 64 LIF neurons and small hidden layers, aiming for an energy-efficient implementation of the BMI application. To achieve an energy efficiency model, they added an additional regularization term to loss function to enforce sparse neuronal activity and the smallest weights in tinyRSNN are pruned. In addition, SMORMS3 optimizer is used and the weight is quantized as float 16 after training. Eventually, they have achieved a mean R^2 score of 0.6978 for bigRSNN and 0.6604 for tinyRSNN.

C. Team Primate Whisperers

[23] from Karlsruhe Institute of Technology (Jann Krausse et al.) won third place in this contest. Inspired by the video of cursor movements of primates, they only focused on a few key points within a reach and applied interpolation to predict the entire route. They implemented a model with a GRU cell followed by an FC layer to determine output, and a temporal convolution layer before it for extracting temporal features. Even more interestingly, they found that the two monkeys have fundamental differences in neural encoding schemes, accuracy decreased when the training data included all three recordings for each primate. Eventually, they have achieved 0.6209 of R^2 score.

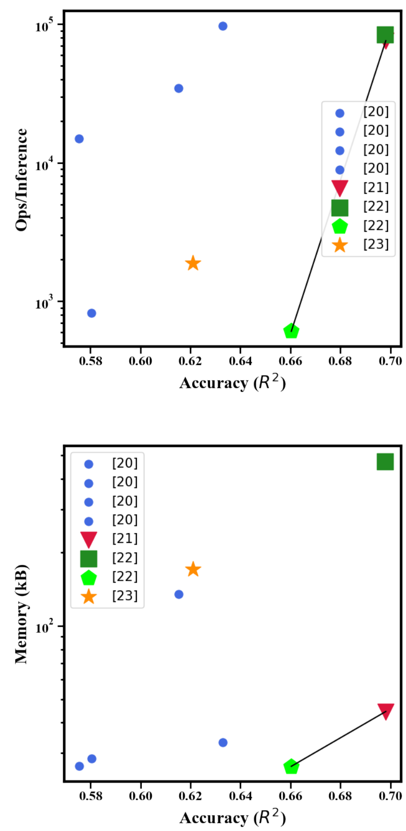

Fig. 2: Pareto plot showing the trade-off between accuracy and computes. The circle makers display the baseline models from [20], while the others represent the results from three teams: a) Compute cost vs. accuracy b) Memory footprint vs. accuracy

VII. Conclusion

A total of 10 teams from 13 universities/organizations/institutions participated in the IEEE BioCAS Grand Challenge to develop algorithms on neural decoding for motor control of non-human primates using an open-access dataset and code harness from [20]. This grand challenge provides an opportunity for the top-3 teams to present their work. We hope that this grand challenge will increase researchers' interest in developing more energy-efficient models and push the boundary for neural decoder design toward next-generation BMI system.

References

[1] "Paralysis statistics - reeve foundation," https://www.christopherreeve.org/living-with-paralysis/stats-about-paralysis.

[2] C. Pandarinath, P. Nuyujukian, C. H. Blabe, B. L. Sorice, J. Saab, F. R. Willett, L. R. Hochberg, K. V. Shenoy, and J. M. Henderson, "High performance communication by people with paralysis using an intracortical brain-computer interface," *elife*, vol. 6, p. e18554, 2017.

[3] C. Libedinsky, R. So, Z. Xu, T. K. Kyar, D. Ho, C. Lim, L. Chan, Y. Chua, L. Yao, J. H. Cheong *et al.*, "Independent mobility achieved through a wireless brain-machine interface," *PLoS One*, vol. 11, no. 11, p. e0165773, 2016.

[4] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, "High-performance neuroprosthetic control by an individual with tetraplegia," *The Lancet*, vol. 381, no. 9866, pp. 557–564, 2013.

[5] C. Cooney, R. Folli, and D. Coyle, "Neurolinguistics research advancing development of a direct-speech brain-computer interface," *IScience*, vol. 8, pp. 103–125, 2018.

[6] I. H. Stevenson and K. P. Kording, "How advances in neural recording affect data analysis," *Nature neuroscience*, vol. 14, no. 2, pp. 139–142, 2011.

[7] A. Nurmikko, "Challenges for large-scale cortical interfaces," *Neuron*, vol. 108, no. 2, pp. 259–269, 2020.

[8] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain–computer interface," *nature*, vol. 442, no. 7099, pp. 195–198, 2006.

[9] D. A. Borton, M. Yin, J. Aceros, and A. Nurmikko, "An implantable wireless neural interface for recording cortical circuit dynamics in moving primates," *Journal of neural engineering*, vol. 10, no. 2, p. 026010, 2013.

[10] H. Miranda, V. Gilja, C. A. Chestek, K. V. Shenoy, and T. H. Meng, "Hermesd: A high-rate long-range wireless transmission system for simultaneous multichannel neural recording applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 4, no. 3, pp. 181–191, 2010.

[11] S. Kim, R. A. Normann, R. Harrison, and F. Solzbacher, "Preliminary study of the thermal impact of a microelectrode array implanted in the brain," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 2986–2989.

[12] T. M. Seese, H. Harasaki, G. M. Saidel, and C. R. Davies, "Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue to chronic heating." *Laboratory investigation; a journal of technical methods and pathology*, vol. 78, no. 12, pp. 1553–1562, 1998.

[13] D. A. Schwarz, M. A. Lebedev, T. L. Hanson, D. F. Dimitrov, G. Lehew, J. Meloy, S. Rajangam, V. Subramanian, P. J. Ifft, Z. Li *et al.*, "Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys," *Nature methods*, vol. 11, no. 6, pp. 670–676, 2014.

[14] W. Zhao, B. Sun, T. Wu, and Z. Yang, "On-chip neural data compression based on compressed sensing with sparse sensing matrices," *IEEE transactions on biomedical circuits and systems*, vol. 12, no. 1, pp. 242–254, 2018.

[15] T. Wu, W. Zhao, E. Keefer, and Z. Yang, "Deep compressive autoencoder for action potential compression in large-scale neural recording," *Journal of neural engineering*, vol. 15, no. 6, p. 066019, 2018.

[16] Y. Chen, E. Yao, and A. Basu, "A 128-channel extreme learning machine-based neural decoder for brain machine interfaces," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 3, pp. 679–692, 2015.

[17] S. Shaikh, R. So, T. Sibindi, C. Libedinsky, and A. Basu, "Towards intelligent intracortical bmi (i ˆ2 bmi): Low-power neuromorphic decoders that outperform kalman filters," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1615–1624, 2019.

[18] M. A. Shaeri, U. Shin, A. Yadav, R. Caramellino, G. Rainer, and M. Shoaran, "33.3 mibmi: A 192/512-channel 2.46 mm^2 miniaturized brain-machine interface chipset enabling 31-class brain-to-text conversion through distinctive neural codes," in *2024 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 67. IEEE, 2024, pp. 546–548.

[19] J. G. Makin, J. E. O'Doherty, Cardoso, M. M. B., and P. N. Sabes, "Superior arm-movement decoding from cortex with a new, unsupervised-learning algorithm," *Journal Neural Engineering*, vol. 15, no. 2, 2018.

[20] J. Yik, S. H. Ahmed, Z. Ahmed, B. Anderson, A. G. Andreou, C. Bartolozzi, A. Basu, D. den Blanken, P. Bogdan, S. Buckley *et al.*, "Neurobench: Advancing neuromorphic computing through collaborative, fair and representative benchmarking," 2023.

[21] Y. Wang, Z. Wang, and S.-C. Liu, "Leveraging recurrent neural networks for predicting motor movements from primate motor cortex neural recordings," in *2024 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2024.

[22] T. Liu, J. Gygax, J. Rossbroich, Y. Chua, S. Zhang, and F. Zenke, "Decoding finger velocity from cortical spike trains with recurrent spiking neural networks," in *2024 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2024.

[23] J. Krausse, A. Vasilache, K. Knobloch, and J. Becker, "Spiking neural networks for low-power intra-cortical brain-machine interfaces," in *2024 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2024.

[24] "Neurotechnology market," https://www.expertmarketresearch.com/reports/neurotechnology-market.

[25] "Bci market," https://www.globenewswire.com/en/news-release/2021/08/12/2279545/0/en/Global-Brain-Computer-Interface-Market-Is-Expected-to-Reach-5-46-Billion-by-2030-Says-AMR.html.

[26] "Neuralink," https://www.neuralink.com/.